

Filling Voids in SRTM Data

Probabilistic Integration of Elevation Data

Ed Wright, Ph.D.

ewright@iet.com, (703)284-0609

Information Extraction & Transport (IET)

1911 N. Ft. Myer Drive, Suite 600

Arlington, VA 22209



Outline

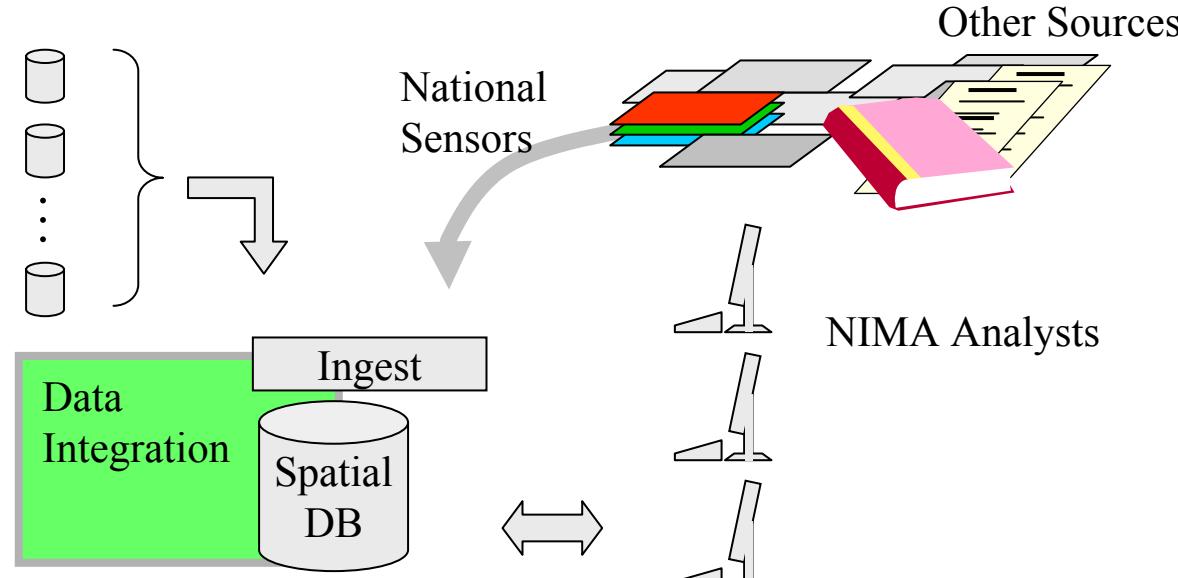


-
- Background & Motivation
 - Bayesian Network Overview
 - Bayesian Networks for Data Integration
 - Graphical Models
 - DEM Integration
 - 2nd Order Uncertainty
 - Ancillary data
 - Summary

Data Integration – NGA SBIR



DFAD,
VMAP,
FFD, ITD,
MSDS, ...



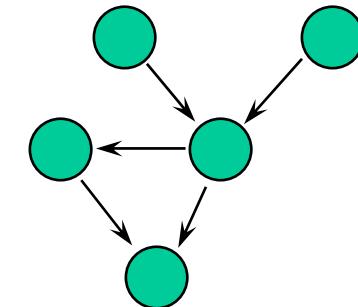
Or
“Data Analysis and Content Integration”

Problem: A future geospatial database will contain conflicting information about spatial objects and their attributes when there are multiple representation from diverse -uncertain and incomplete - sources over time.

Solution: maintain a consistent set of hypotheses about the existence, type, and attributes of spatial objects, and the relationships between them. The hypotheses are instantiated and probabilistically evaluated based on evidence from diverse sources.

Relatively new AI technique only now being applied to real problems

- Graphical model
- Uncertain variables
- Probability distributions that relate variables
- Propagation algorithm
 - Prior probabilities
 - Propagation of evidence



Much more powerful and flexible than a traditional Rule Base, and rule based expert system

Exploit existing data or expert knowledge to get probability distributions

Works with discrete or continuous variables

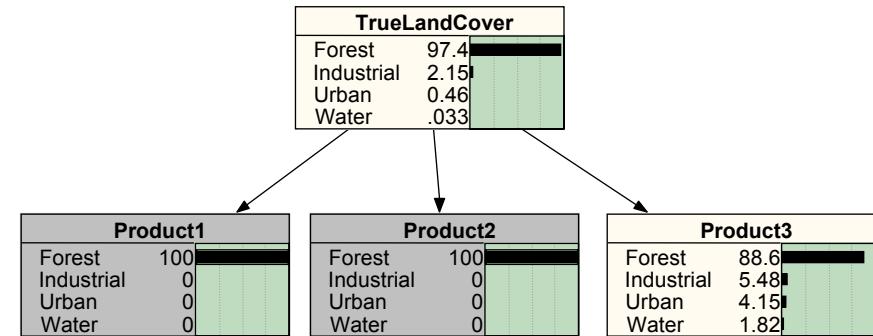
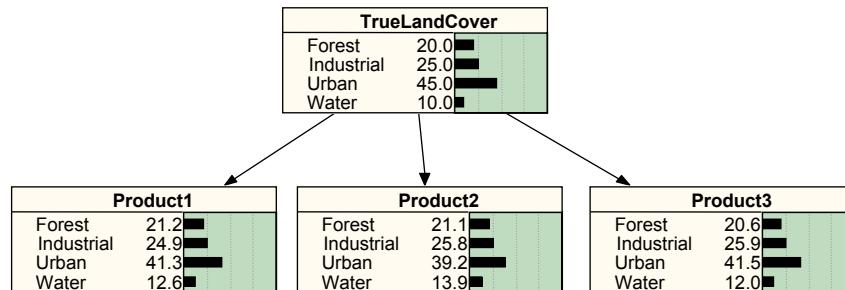
Successful applications:

- Medical diagnosis, Maintenance diagnosis, Planning under uncertainty, Uncertain reasoning, Uncertainty in AI

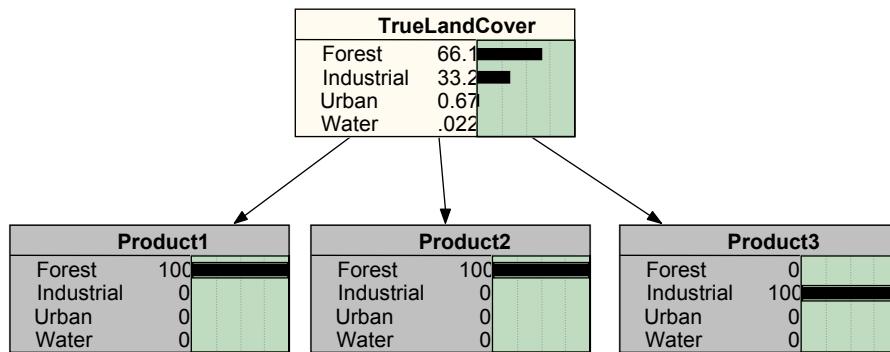
Data Integration



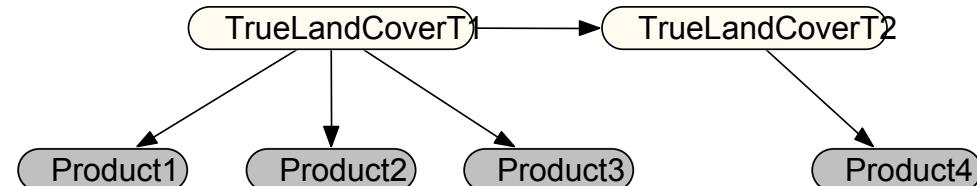
- Bayesian Network for Integration



Incomplete information



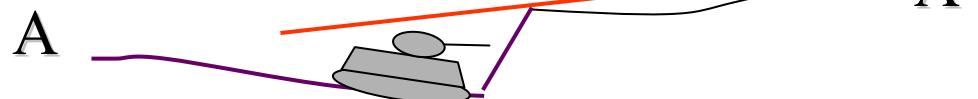
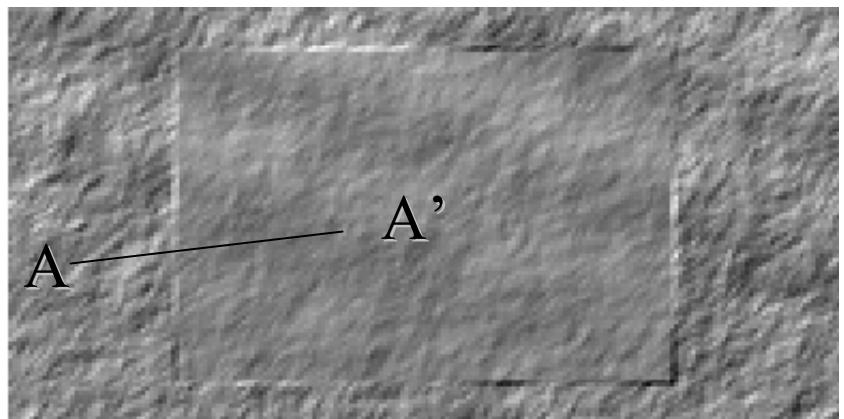
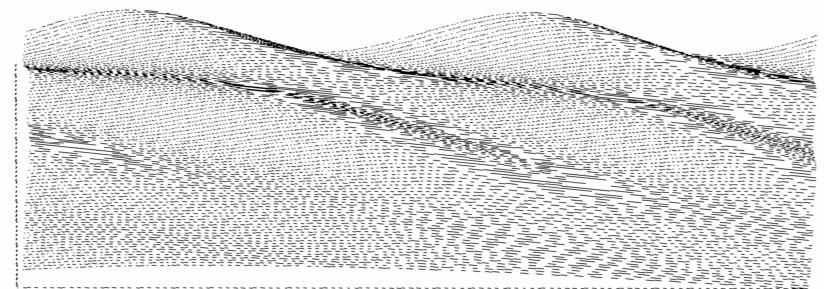
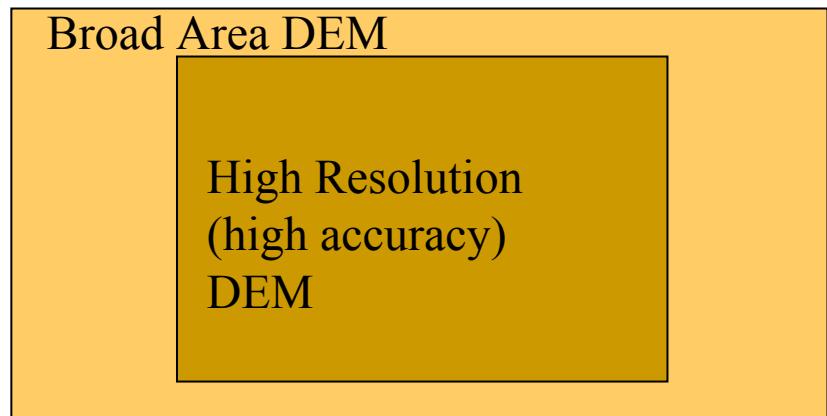
Conflicting information



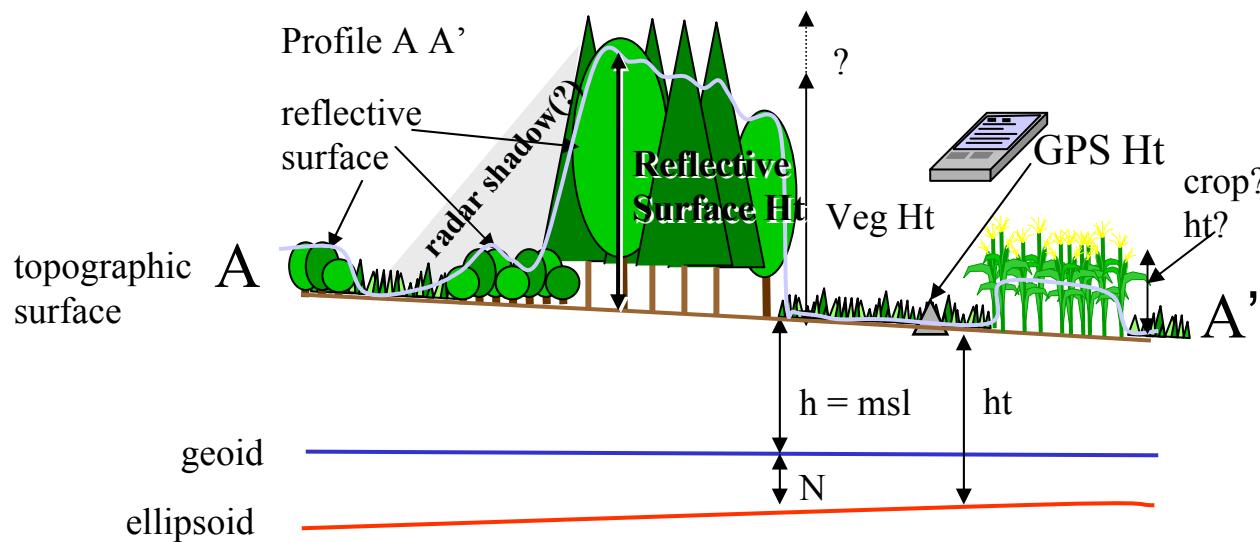
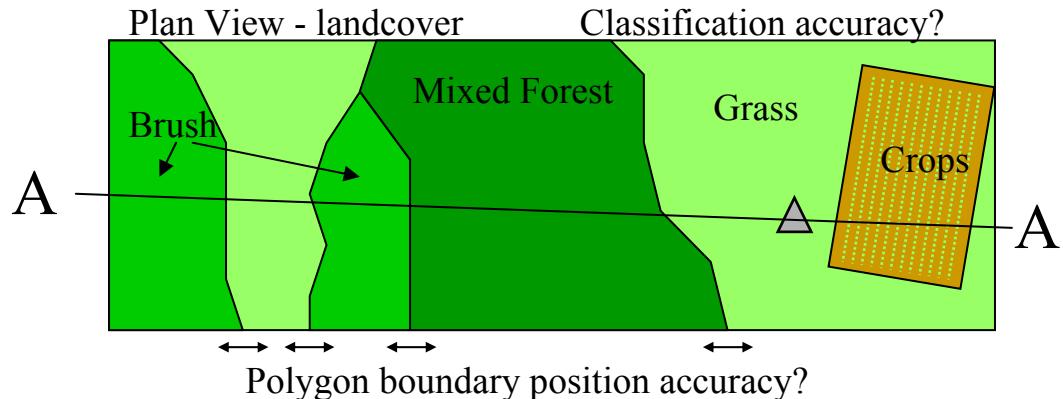
Modeling temporal changes

Data Integration

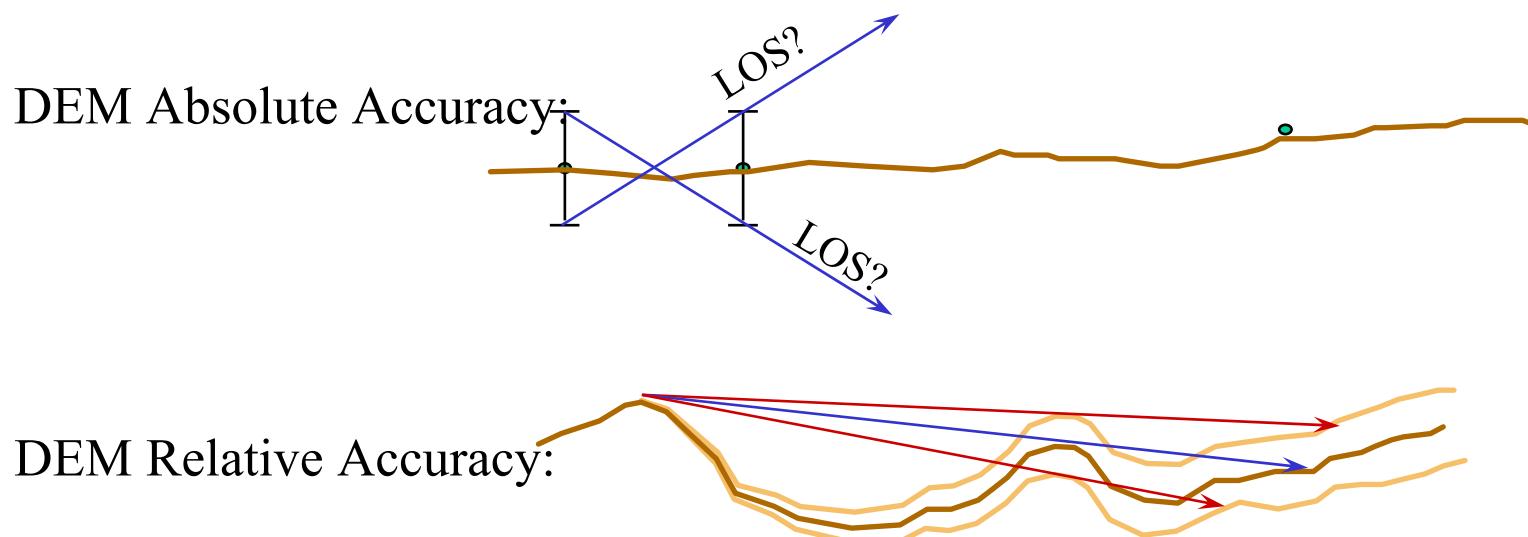
- How to combine Multiple sources of elevation data
 - Use highest resolution?
 - Use most accurate?
 - Use Most recent?
- How good is the result?



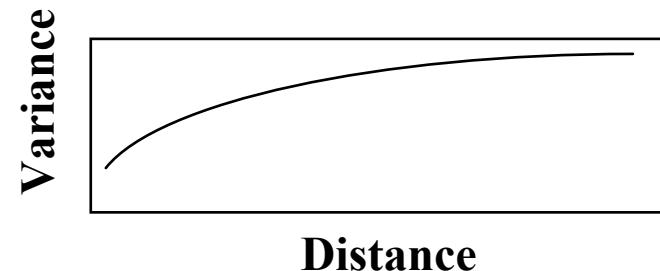
Challenges in Elevation Integration



Accuracy of LOS?



Geostatistics: Variogram is used to estimate correlation based on the distances between points.



Probability Models for Spatial Data



- Traditional Statistical Tools of the mapping sciences:

- Least Square Adjustment (LSA)
- Error Propagation

$$Y = F(X)$$

$$\Sigma_{YY} = G \Sigma_{XX} G^T$$

$$G = \frac{\partial F(X)}{\partial X}$$

- Geostatistics
- Kriging

$$\lambda(s) = C^{-1}c(s)$$

^

$$U(s) = \lambda^t(s)U = c^t(s)C^{-1}U$$

$$\text{Kriging Variance} = \sigma^2 - c^t(s)C^{-1}c(s)$$

$$F(L_a, X_a) = 0$$

$$BV + AX + W = 0$$

$$M = BP^{-1}B^T$$

$$N = A^T M^{-1} A$$

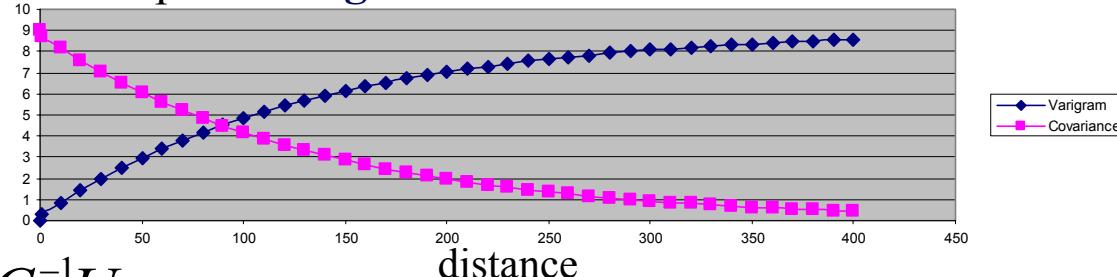
$$U = A^T M^{-1} W$$

$$\hat{X} = -N^{-1}U$$

$$K = -M^{-1}(A\hat{X} + W)$$

$$V = P^{-1}B^T K$$

Example Variogram and Covariance functions



$c(s)$, and C , are each computed from the Covariance function

“Standard Error Theory” | Bayesian Networks



- Bayesian Networks can be used for the same kinds problems as standard error theory

	<u>Standard Error Theory</u>	<u>Bayesian Networks</u>
Forward Solution:	Error Propagation	Error Propagation
Inverse Solution:	“Learn” (adjusted) Parameters	Learn Models / parameters
Data Quality:	Variance Covariance Matrix	Probability Distributions
Visualize Uncertainty:	Error Ellipse (Ellipsoid)	Histograms
Spatial Correlation:	Kriging, LS Collocation	<i>Graphical Models</i>

Graphical Models

- Undirected Arcs - Spatial Correlation
- In general no exact solution
- Markov Chain Monte Carlo (MCMC) - very large computation

Data Integration: DEMs (Spatial Correlation)

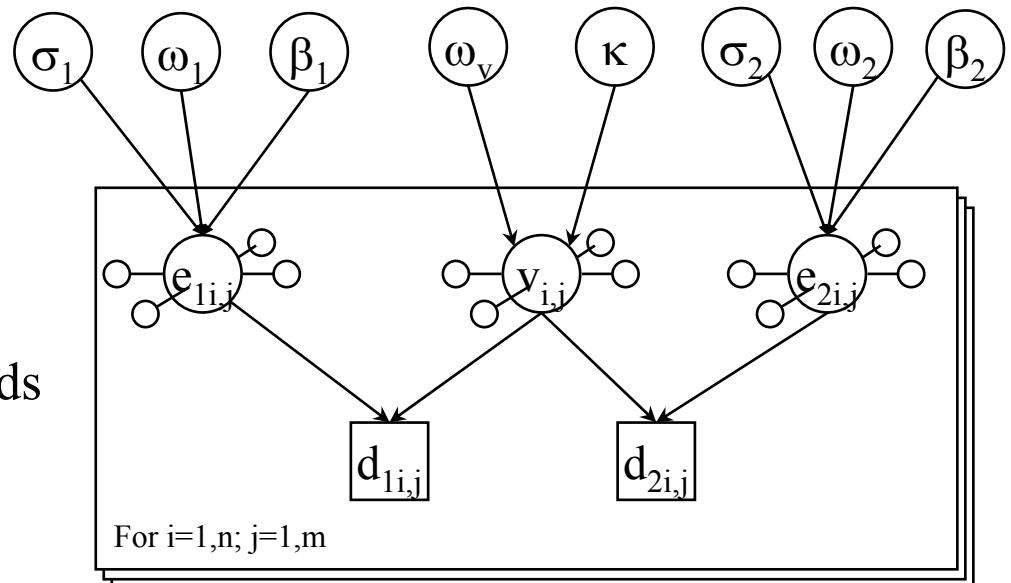
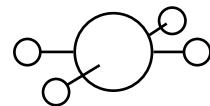


Elevations on two DEMs

$$\text{DEM}_1: d_{1i,j} = v_{i,j} + e_{1i,j}$$

$$\text{DEM}_2: d_{2i,j} = v_{i,j} + e_{2i,j}$$

e_1, e_2, v : Markov Random Fields

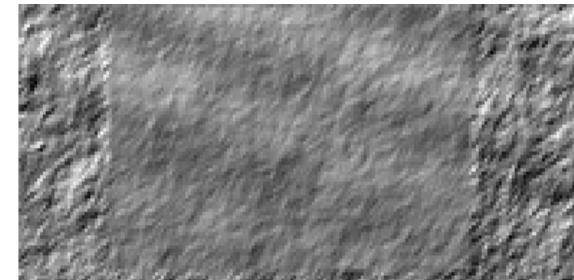
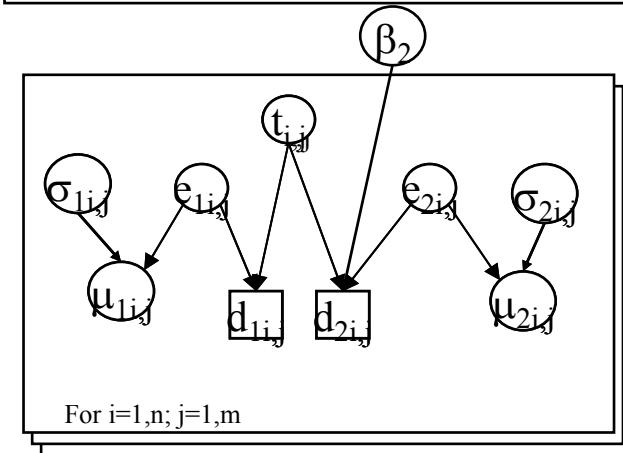
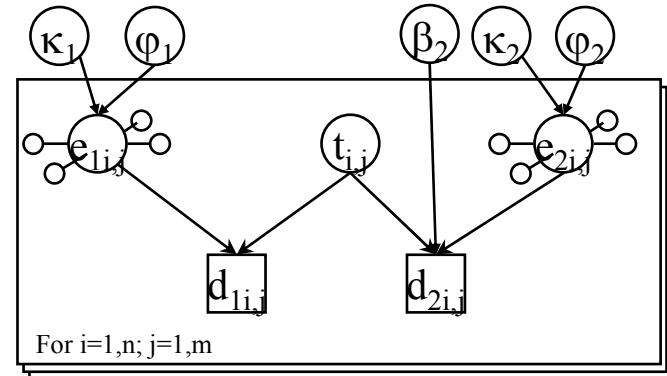


Combining Elevation Surfaces



- Extremely large Model
- Graphical Model: not Bayesian Network
 - Use MCMC
 - $P(e_{1i,j} | \text{Markov Blanket})?$
- Use simple kriging for influence of neighbors:
 - $P(e_{1i,j} | \text{neighbors}) = N[\mu, \sigma^2]$
 - $\mu = \sum \lambda e_n$, λ = Kriging weights
 - $\sigma^2 = F(\lambda)$, Kriging variance
- One iteration => probability distributions for each post ($e_{1i,j}$, $e_{2i,j}$, $t_{i,j}$)
 - calc kriging mean(s), std dev(s) => instantiate nodes (for each random field)
 - use observed data (d) to instantiate nodes
 - Least Squares Adjustment for distribution, and sample from it
- Iterate until convergence - Result:

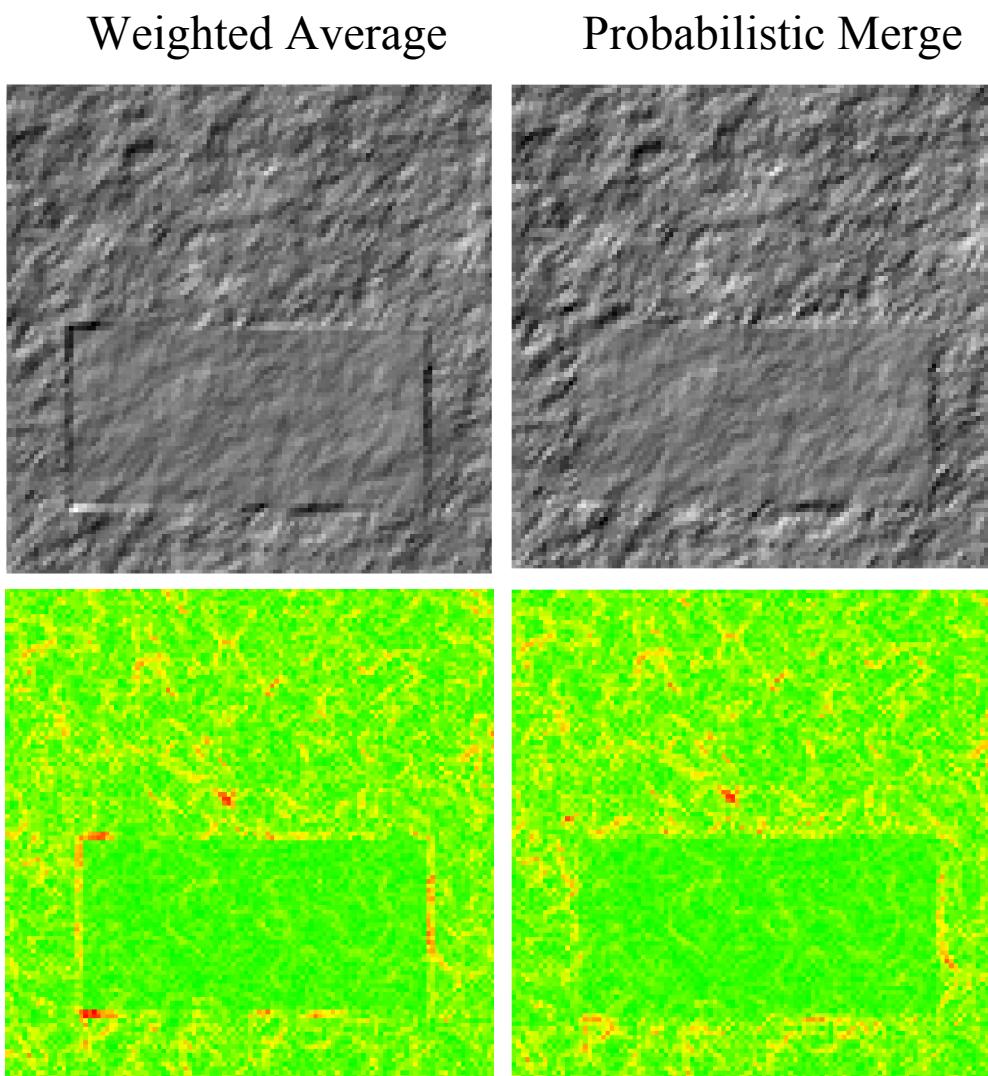
**Easily parallelizable,
extensible to categorical data**



Digital Elevation Model (DEM) Integration: simulated data



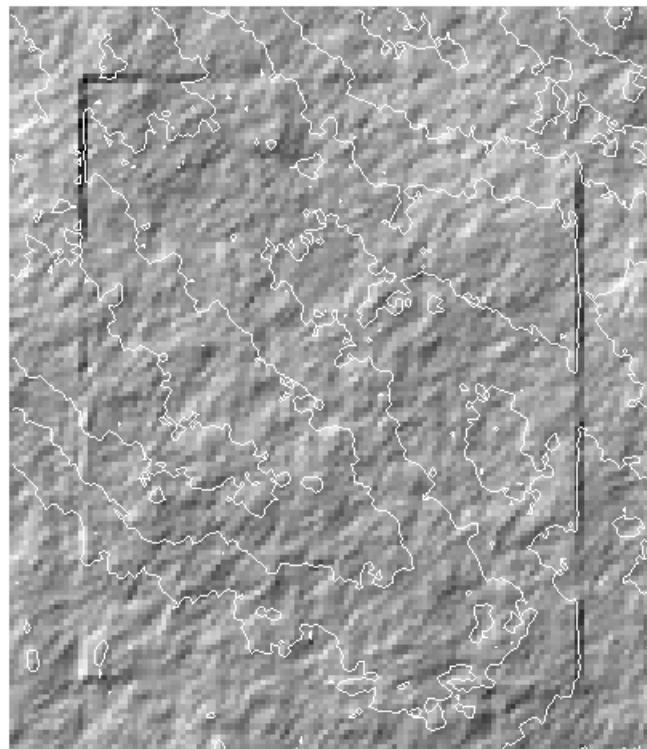
Analytic
Hill
Shading



Digital Elevation Model (DEM) Integration: simulated data

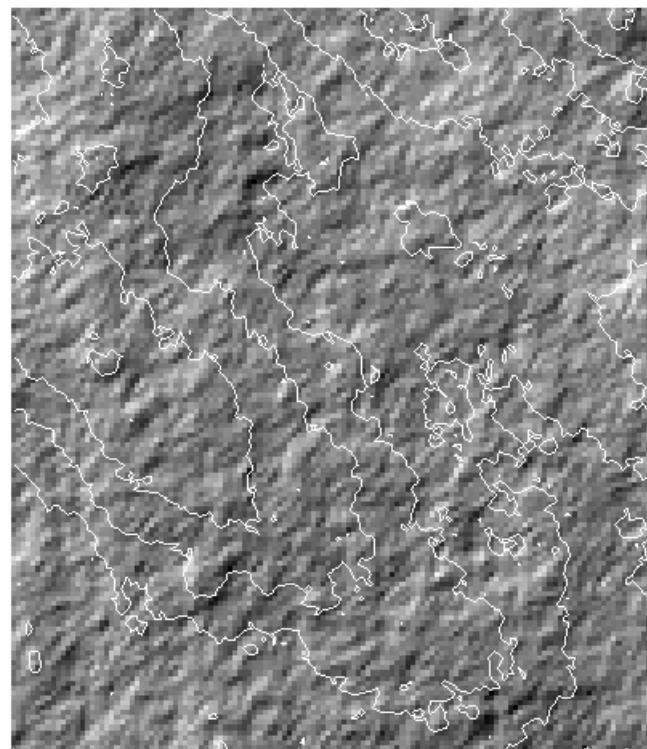


Substitution



0.07
0.11
0.15
0.20
0.24
0.29
0.33
0.37
0.42
0.46
0.50
0.55
0.59
0.63
0.68
0.72
0.77

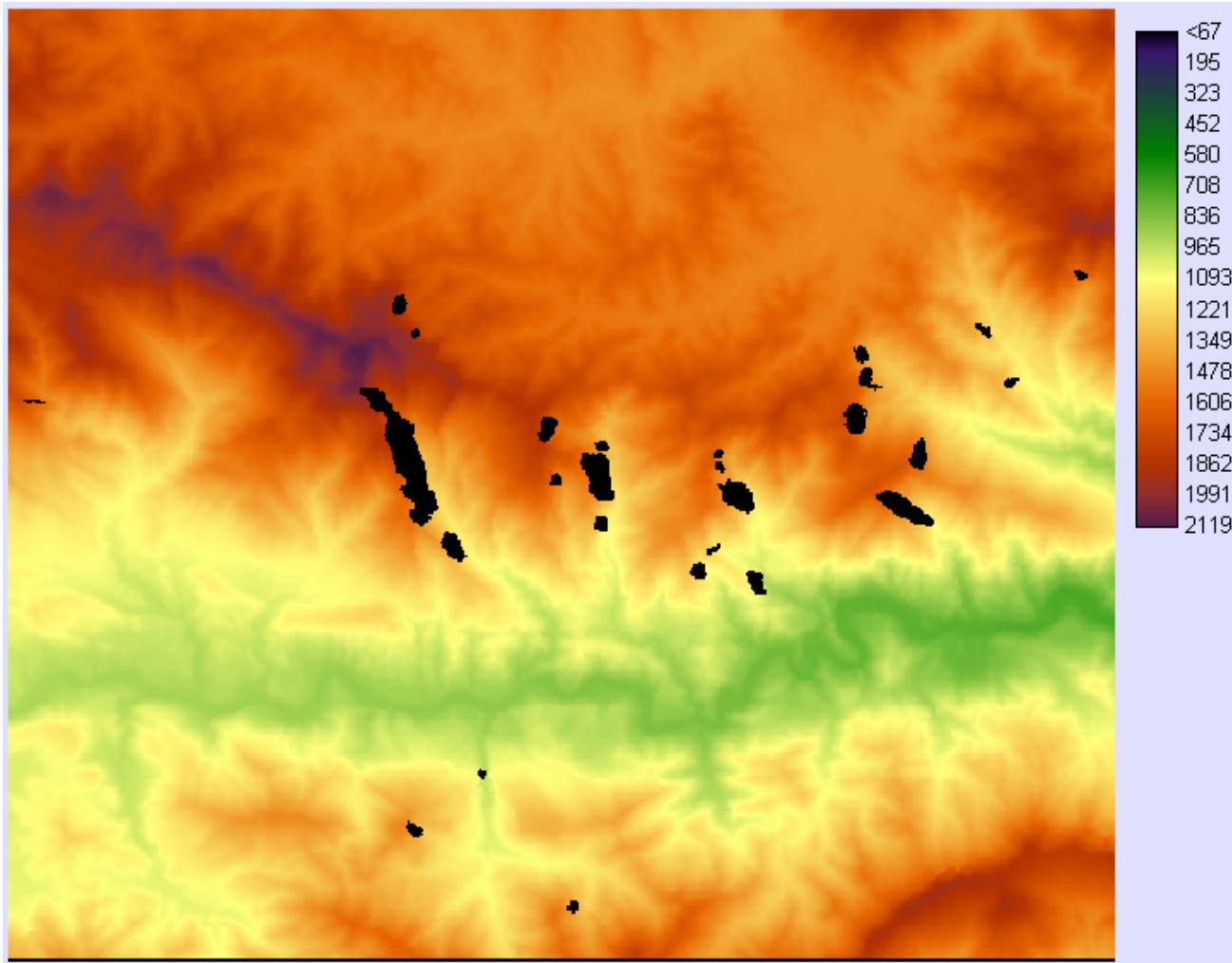
Probabilistic merge



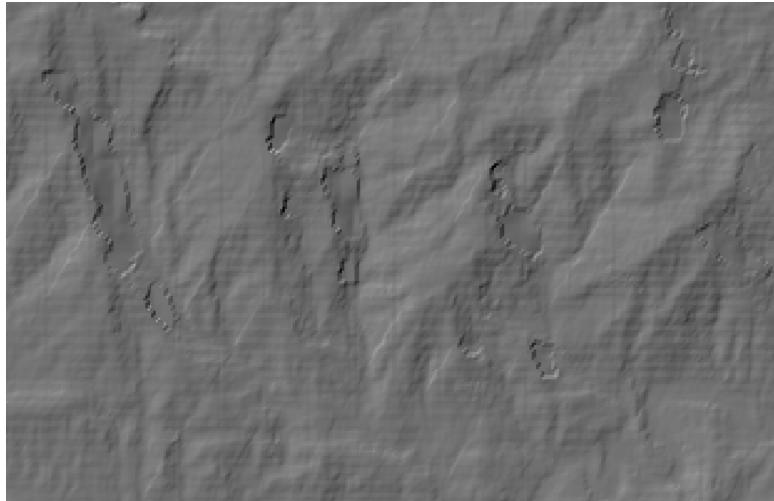
0.21
0.25
0.28
0.32
0.35
0.39
0.42
0.45
0.49
0.52
0.56
0.59
0.63
0.66
0.69
0.73
0.76

Analytical Hill Shading

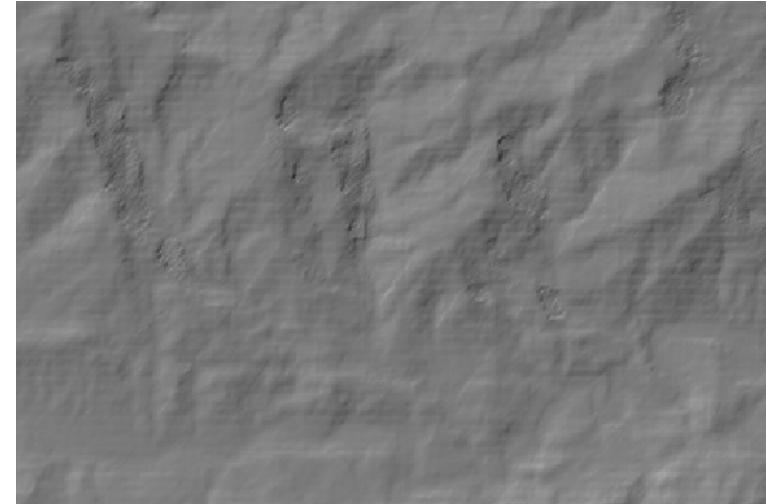
SRTM (subset) with voids



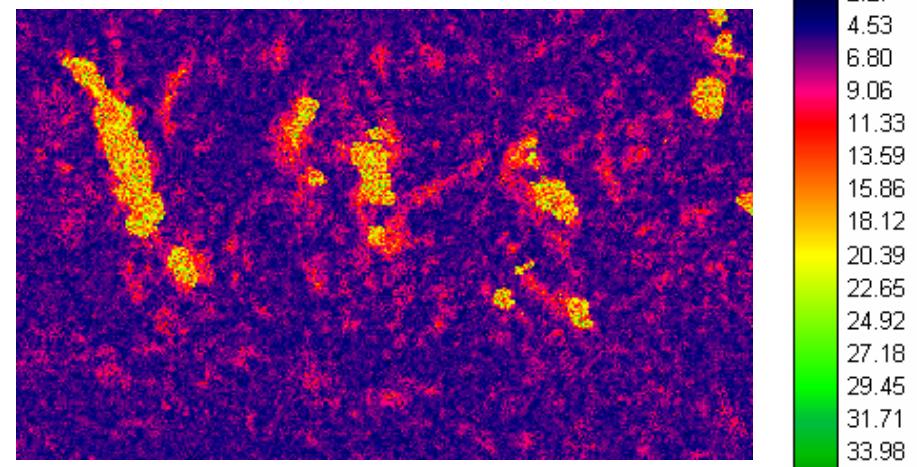
Digital Elevation Model (DEM) Integration: SRTM with DTED 1



Replace: Level 1 in SRTM Holes



Merge: Level 1 with SRTM



Estimated accuracy: Level 1
merged with SRTM

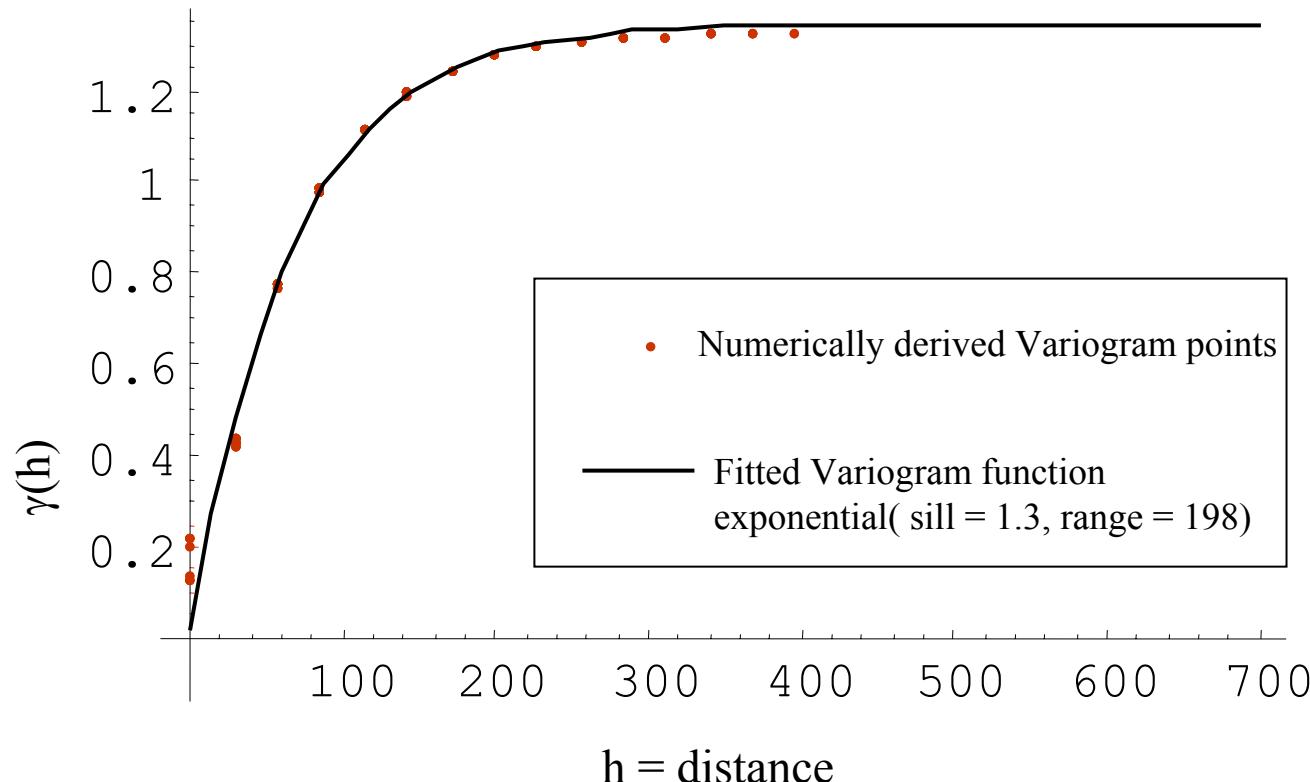
Spatial Accuracy of Integrated DEMs



Analytically propagate the accuracy of a grid of points

- Accuracy defined by the input DEMs

Empirical fit of a variogram to the results



Where Does the Data Quality Information come from?



- Data Quality Meta Data
 - Relative accuracy! As a function of distance
- Infer from Production Data
- Propagate through transformations
 - Resampling introduces correlation
- Second Order Uncertainty

Error - Uncertainty



- Common use in the GIS community
 - Error / accuracy
 - Known error model, with known parameters
 - Uncertainty
 - Unknown error model, or unknown parameters
- Statistics
 - Error \Leftrightarrow uncertainty
 - 2nd Order Uncertainty \Leftrightarrow uncertainty in the uncertainty

2nd Order Accuracy DEM Integration - Results



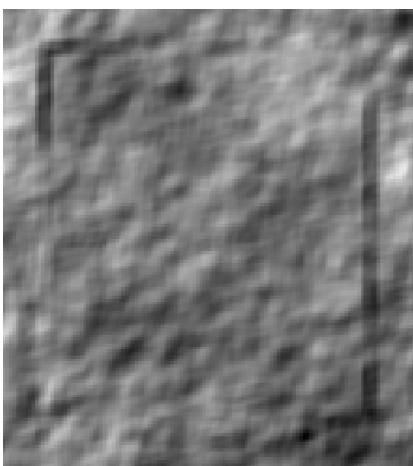
- 2 data sets simulated from an analytic surface, known accuracy
- For the integration, the accuracy of one data set is assumed not known
- Actual error: $C_{10} = 4.0$, meters and range, $a_1 = 400$ meters

Burn in	Total Iterations	Starting Estimates		Sill C01			Range a1	
		C01	a1	Mean	Std Dev	Mean	Std Dev	
100	200	12	600	4.8	1.29	524	159.64	
400	600	12	600	4.7	0.92	515	106.12	
400	800	12	600	4.7	1.06	517	124.64	
400	800	2	600	4.7	1.05	514	125.58	
400	800	4	400	4.8	1.14	522	134.71	
400	800	8	200	4.7	1.03	515	124.60	
400	800	6	600	4.9	1.10	536	132.87	
400	800	6	100	4.8	1.05	520	122.49	

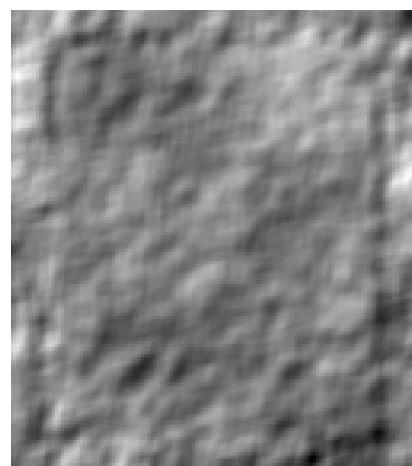
2nd Order Accuracy DEM Integration - Results



Analytic
Hill
Shading

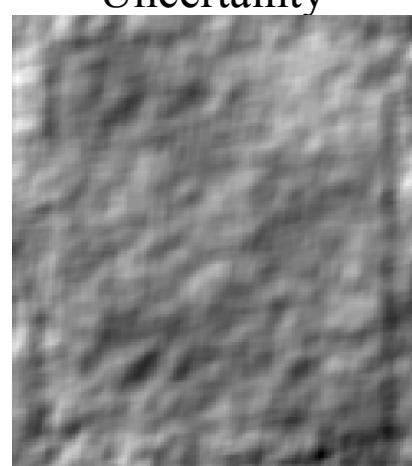


Weighted Average

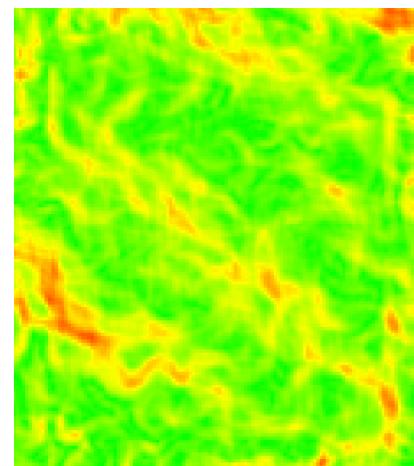
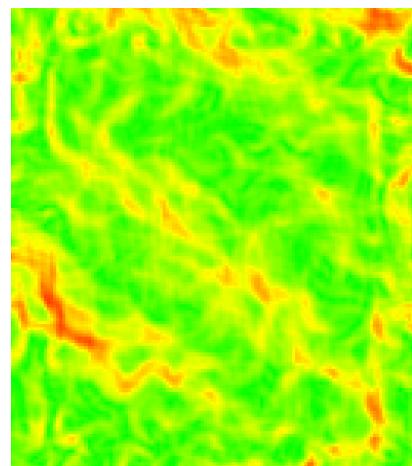
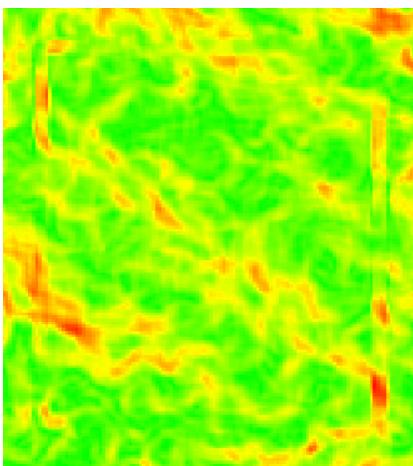


Probabilistic Merge

Probabilistic Merge
with Second Order
Uncertainty



Slope



Data Integration: DEMs (*a priori* Feature data)

Elevations on two DEMs

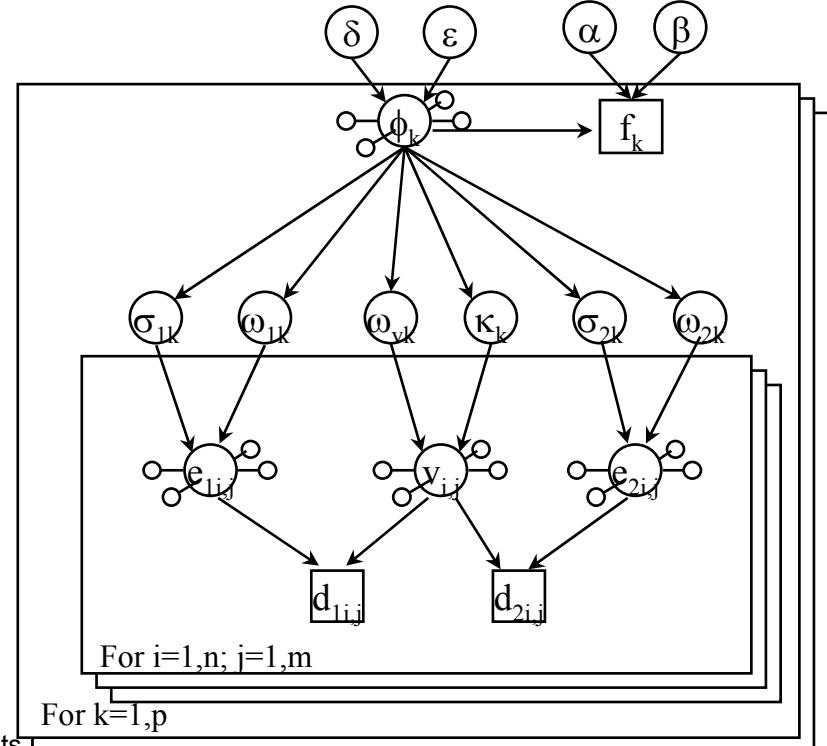
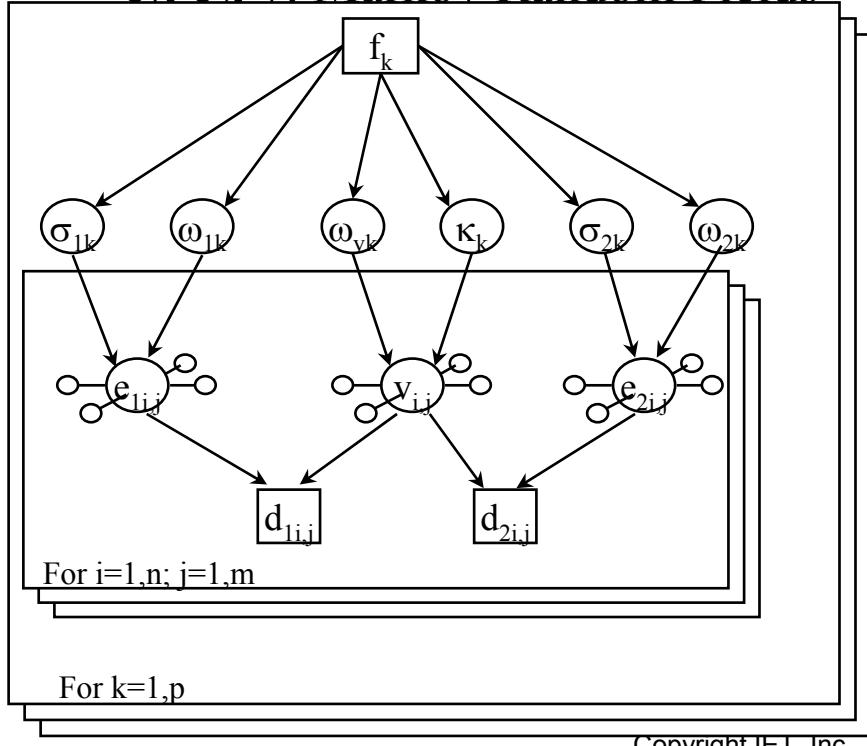
$$\text{DEM}_1: d_{1i,j} = v_{i,j} + e_{1i,j}$$

f_k is feature type (slope, vegetation, ...)

$$\text{DEM}_2: d_{2i,j} = v_{i,j} + e_{2i,j}$$

can also extend to include uncertainty in feature types

e_1, e_2, v : Markov Random Fields



Integration of Elevation Surfaces



- Exploit prior estimate of accuracy of elevation products
- Estimate accuracy of integrated result
 - Supports evaluation of fitness for use - Probabilistic Line of Sight
- 2nd Order Uncertainty: For the case when accuracy of an input is unknown
 - Estimated accuracy of result
 - Estimated accuracy of original
- Correlation introduced by resampling
- Potential:
 - Effects of alternate error models - based on land cover (slope, vegetation)
 - Effects of alternate surfaces: Example - reflective surface vs bare Earth